MicrobesOnline: an integrated portal for comparative functional genomics

Marcin P. Joachimiak[1,2], Y. Wayne Huang[1,2], Katherine H. Huang[1,2], Eric J. Alm[1,3], Dylan Chivian[1,2], Paramvir S. Dehal[1,2], Keith Keller[1,4], Morgan N. Price[1,2], Adam P. Arkin[1,2,4,5]

[1]Virtual Institute for Microbial Stress and Survival, http://vimss.lbl.gov; [2]Lawrence Berkeley National Laboratory, Berkeley, CA, 94720;
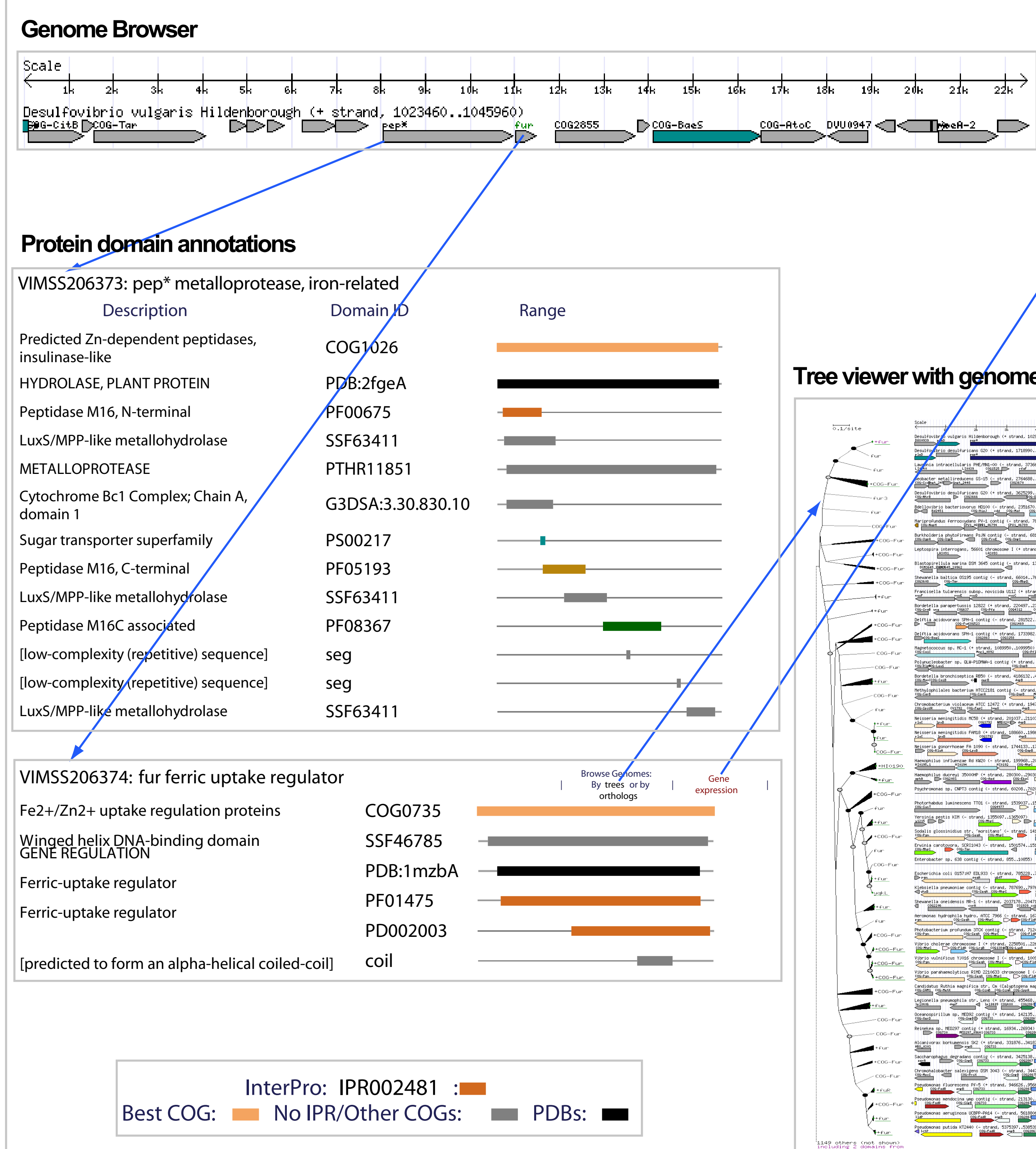[3]Department of Biological Engineering, MIT, Cambridge, MA, 02139; [4]University of California, Berkeley, CA, 94720

DOE GENOMICS:GTL — ACCELERATING DISCOVERY FOR ENERGY AND ENVIRONMENT — OFFICE OF SCIENCE, U.S. DEPARTMENT OF ENERGY

VIMSS Virtual Institute for Microbial Stress and Survival | MIT Massachusetts Institute of Technology | MISSOURI | UNIVERSITY OF WASHINGTON | MONTANA STATE UNIVERSITY | Southern Illinois University Carbondale | OU | Sandia National Laboratories | OAK RIDGE NATIONAL LABORATORY Managed by UT Battelle for the Department of Energy | Berkeley University of California | **http://vimss.lbl.gov/**

Introduction

VIMSS has developed a Web portal, along with computational analysis and an underlying database, for comparative functional genomics of bacteria and archaea. MicrobesOnline (http://www.microbesonline.org) has been enabling comparative genome analysis since 2003. The portal currently includes 702 complete microbial genomes (662 bacterial, 40 archaeal, 1770 viruses, 268 plasmids) and offers a suite of analysis and tools including: comprehensive gene family and domain annotations, information on three dimensional structure representatives and similarities, an interactive gene expression heatmap browser along with functionality for gene expression profile searches, a multi-species genome browser, operon and regulon predictions, a combined gene and species phylogeny browser, a gene ontology browser, a metabolic pathway browser, a workbench for sequence analysis (including sequence motif detection, motif searches, sequence alignment and phylogeny reconstruction), and capabilities for community annotation of genomes.

Genome Browser

Protein domain annotations

VIMSS206373: pep* metalloprotease, iron-related

Description	Domain ID	Range
Predicted Zn-dependent peptidases, insulinase-like	COG1026	
HYDROLASE, PLANT PROTEIN	PDB:2fgeA	
Peptidase M16, N-terminal	PF00675	
LuxS/MPP-like metallohydrolase	SSF63411	
METALLOPROTEASE	PTHR11851	
Cytochrome Bc1 Complex; Chain A, domain 1	G3DSA:3.30.830.10	
Sugar transporter superfamily	PS00217	
Peptidase M16, C-terminal	PF05193	
LuxS/MPP-like metallohydrolase	SSF63411	
Peptidase M16C associated	PF08367	
[low-complexity (repetitive) sequence]	seg	
[low-complexity (repetitive) sequence]	seg	
LuxS/MPP-like metallohydrolase	SSF63411	

VIMSS206374: fur ferric uptake regulator

Fe2+/Zn2+ uptake regulation proteins	COG0735	
Winged helix DNA-binding domain GENE REGULATION	SSF46785	
Ferric-uptake regulator	PDB:1mzbA	
Ferric-uptake regulator	PF01475	
	PD002003	
[predicted to form an alpha-helical coiled-coil]	coil	

InterPro: IPR002481 : ■
Best COG: ■ No IPR/Other COGs: ■ PDBs: ■

Tree viewer with genome browser

Browse Genomes: By trees or by orthologs | Gene expression

VIMSS integrates functional genomic data and provides novel web-based viewing and analysis tools for gene expression microarray compendia, metabolic pathways, gene function and protein domain annotations, regulation (including motif searches, regulon predictions, and integration with RegTransBase). Among the most recently added features are:

* Web-based viewer for microarray data from GTL and other projects mostly for the following organisms:
  - Desulfovibrio vulgaris Hildenborough
  - Bacillus subtilis
  - Helicobacter pylori 26695
  - Shewanella oneidensis MR1
  - Escherichia coli K12
  - Salmonella enterica sv Typhimurium
  - Geobacter metallireducens GS-15
* Interactive heatmaps for gene carts and operons.
* Search for genes whose expression is similar to the query gene(s).
* Phylogenetic trees for practically all gene families.
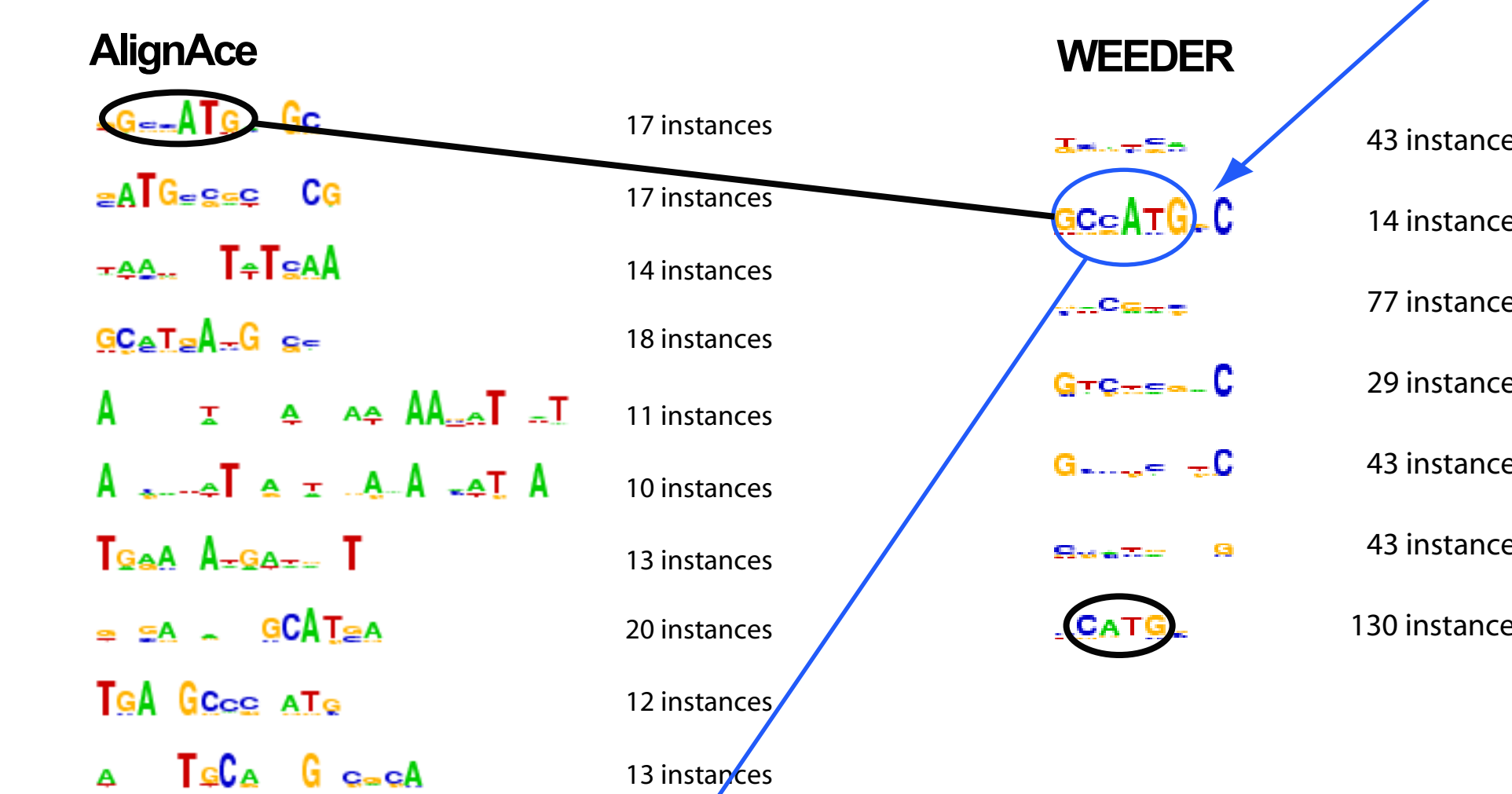* Comprehensive protein sequence domain annotations

Gene, operon, and gene set expression data heatmaps

Currently MicrobesOnline provide interactive gene expression data functionality for *D. vulgaris* Hildenborough, *E. coli* K12, and *S. oneidensis* MR1. Additional microarray data for these and other species from the ESPP2 project as well as other public resources are being incorporated into the MicrobesOnline database and made publicly available. The interactive heatmaps include popups and links detailing gene annotations and experiment meta-data.

Interactive gene expression heatmap
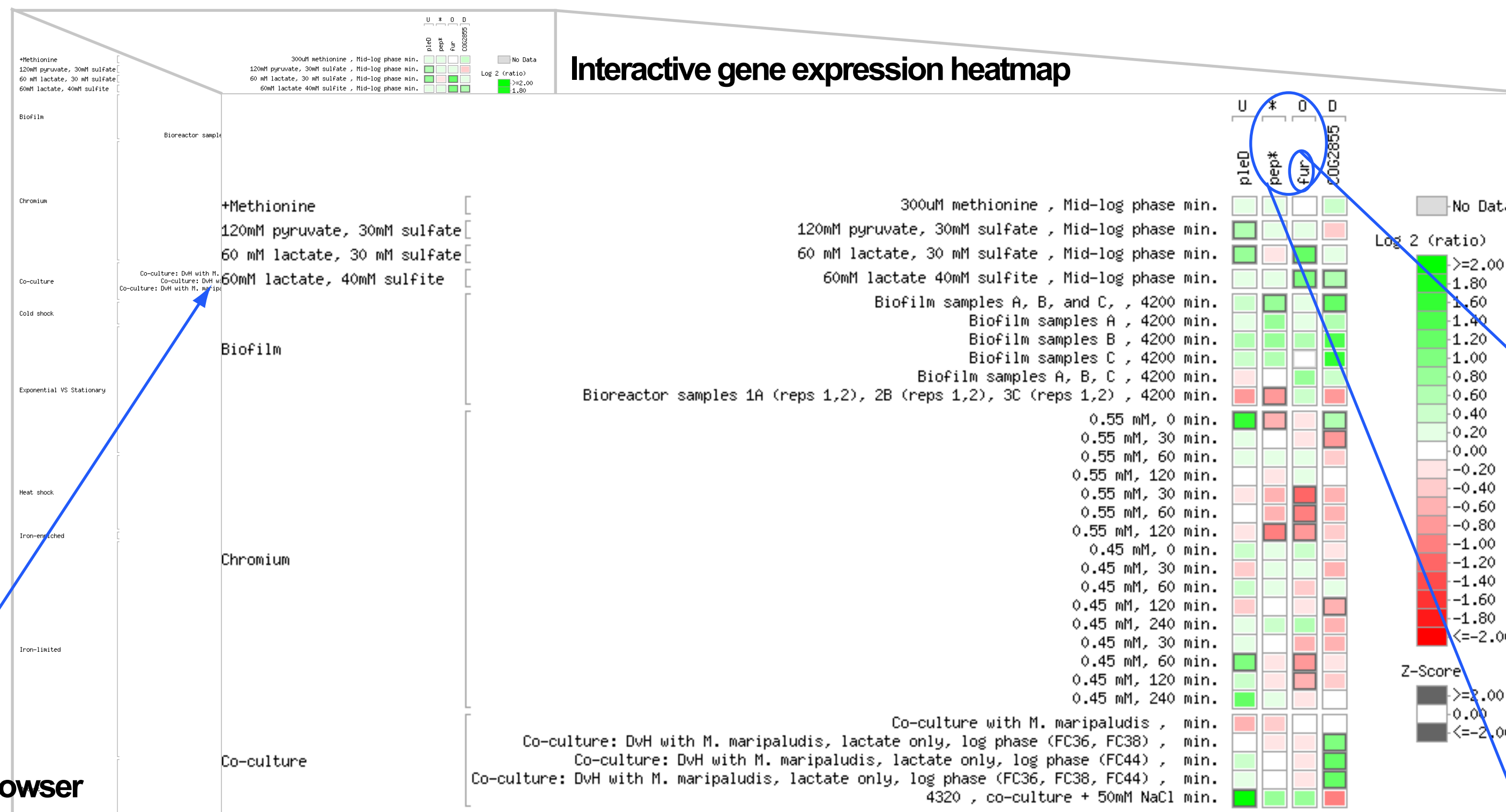
Regulatory sequence motif detection and scanning

MicrobesOnline includes a suite of sequence motif detection tools. Starting from user specified genes in a user's cart, the motif detection interface allows creating custom motif search jobs with the AlignACE, MEME, or WEEDER programs. In this example, based on searching for motifs in the top 15 matches to the fur operon expression profile search (Pearson correlation > 0.55), two programs agreed on a motif of which 14-17 instances were found.

AlignAce

17 instances
17 instances
14 instances
18 instances
11 instances
10 instances
13 instances
20 instances
12 instances
13 instances

WEEDER

43 instances
14 instances
77 instances
29 instances
43 instances
43 instances
130 instances

The motif search results are displayed as a list of sequence logos with statistics for each detected sequence motif match. The identified motifs can subsequently be used to search genomes, taxonomic groups of genomes, or genes in carts using the programs scanACE, MAST, and patser. This process can be iterated by using the results of a motif scan as the set of sequences for subsequent motif detection. Based on the motif identified above, a large number of candidate matches were identified for further analysis.

Score	E-value	Sequence	Gene	Name	Offset	Strand	Gene Description
9.62	1.53e-05	GCCATGAC	DVU0010	COG-DctM	84	-	TRAP transporter, DctQ family
9.62	1.53e-05	GCCATGAC	DVU0033	COG-PncA	-85	+	isochorismatase family protein
9.62	1.53e-05	GCCATGAC	DVU0044	fliP	6	-	flagellar biosynthetic protein fliP
9.62	1.53e-05	GCCATGAC					
9.62	1.53e-05	GCCATGAC					
9.62	1.53e-05	GCCATGAC					

Interactive gene expression profile searches

As of recently MicrobesOnline also provides interactive gene expression profile searches for *D. vulgaris* Hildenborough, *E. coli* K12, and *S. oneidensis* MR1. Dynamic profile searches are available for single genes, genes in an operon, or genes in carts. In the latter two modes, the query profiles are constructed from the mean expression of genes in the operon or cart. Using sets of selected genes can improve signal when mining gene expression data.

gene expression profile

operon expression profile
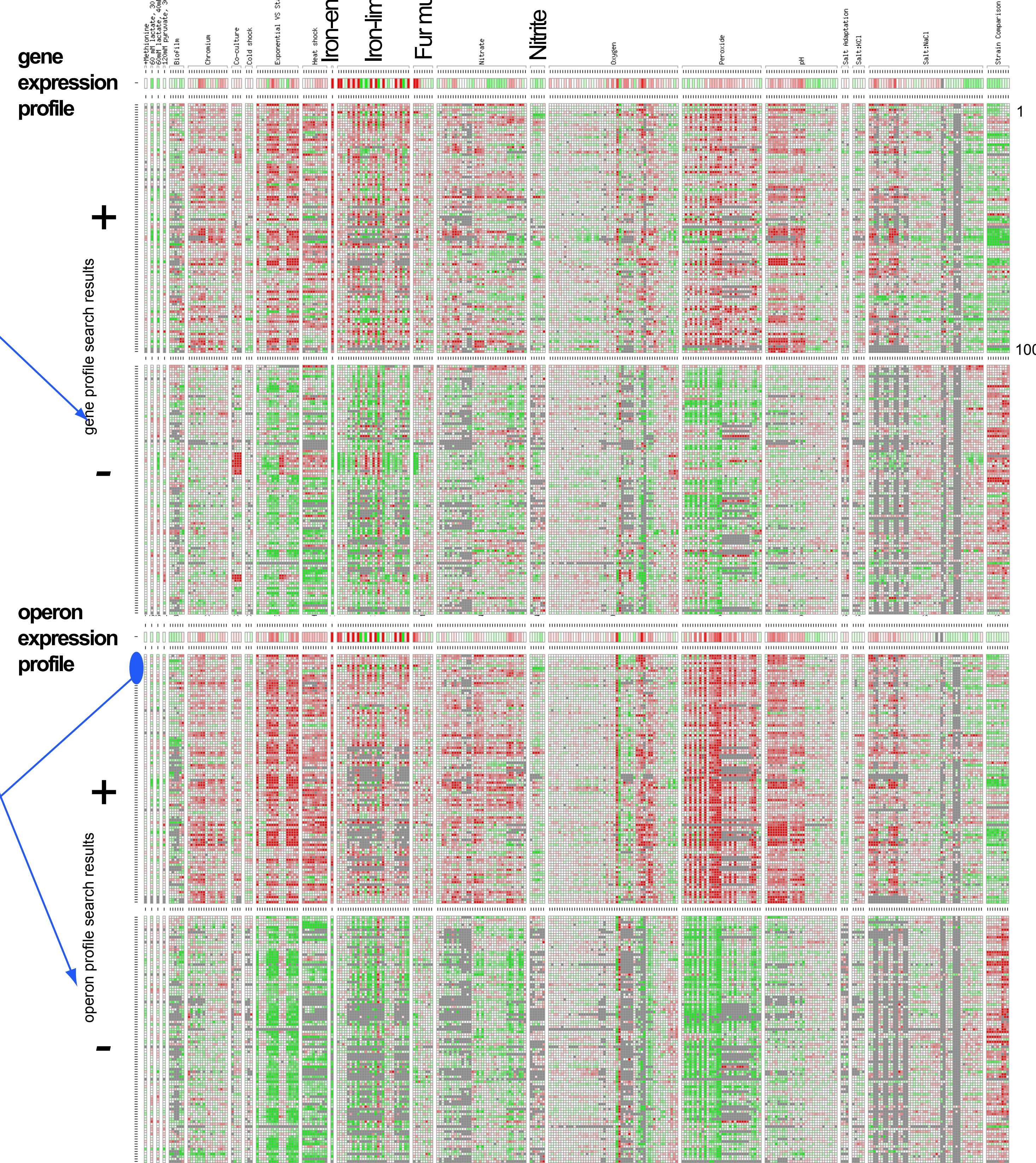
Conclusions

These new data compendium-wide functionalities allow to observe patterns in gene expression changes across multiple conditions and genes, and to search for similarities to these patterns. The gene cart feature allows to apply different analysis to the same set of genes thereby integrating multiple layers of information. The information integration and analysis performed by VIMSS serves not only to generate insights into the stress responses and their regulation in these microorganisms, but also to document experiments, allow contextual access to experimental data, and facilitate the planning of future experiments. VIMSS also is incorporating into MicrobesOnline publicly available functional genomics data from published research, so as to centralize and synergize data on and analysis of microbial physiology and ecology in a unified comparative functional genomic framework.

References

*Alm EJ, Huang KH, Price MN, Koche RP, Keller K, Dubchak IL, Arkin AP. The MicrobesOnline Web site for comparative genomics. Genome Res. 2005 Jul;15(7):1015-22.
*Bender KS, Yen HB, Hemme CL, Yang Z, He Z, He Q, Zhou J, Huang KH, Alm EJ, Hazen TC, Arkin AP, Wall J Analysis of a Ferric Uptake Regulator (Fur) Mutant of Desulfovibrio vulgaris. AEM 2007 Sep;73(17):5389–5400.
*He Q, Huang KH, He Z, Alm EJ, Fields MW, Hazen TC, Arkin AP, Wall J, Zhou J Energetic Consequences of Nitrite Stress in Desulfovibrio vulgaris Hildenborough, Inferred from Global Transcriptional Analysis. AEM 2006 June; 72(6): 4370–4381.

Acknowledgment